Created by Neil Yetz & Gemma Wallace

Module 11 Lab Answer Key: Core Concepts in Analysis of Variance
PSY 652 Research Methods
Nov 13, 2019

```
1   ---
2   title: "Module 11 lab"
3   subtitle: Gemma Wallace & Neil Yetz
4   output:
5     html_document:
6       df_print: paged
7       toc: yes
8     html_notebook:
9       toc: yes
10  ---
11
12  # Load libraries
13  ```{r, message = FALSE}
14  library(tidyverse)
15  library(psych)
16  ```
17
18  # Import data
19  ```{r,message=FALSE}
20  stress <- read_csv("StressReduction.csv")
21  ```
22
23  # Factor the grouping variable
24  ```{r}
25  stress <- stress %>% mutate(ProgramCode.f = factor(ProgramCode, levels = c(0,1,2,3),labels = c("None", "Healthy Habits", "Get
    Movin", "Recharge")))
26  ```
27
28
29  # Get descriptives
30  ```{r, warning = FALSE}
31  describe(stress$StressLevel)
32  mean(stress$StressLevel)
33  aggregate(x=stress$StressLevel, by=list(stress$ProgramCode.f), FUN=mean)
34  ```
```

```
34  ```
```

data.frame 1 x 13	R Console	data.frame 4 x 2

Group.1 <fctr>	x <dbl>
None	6.333333
Healthy Habits	5.833333
Get Movin	4.533333
Recharge	5.200000

4 rows

35 The mean stress level for each of the four groups are all different from the grand mean (5.475), indicating that stress level
    varied across groups. The mean stress level for the "none" (i.e., no treatment) group was highest (6.333), while the mean
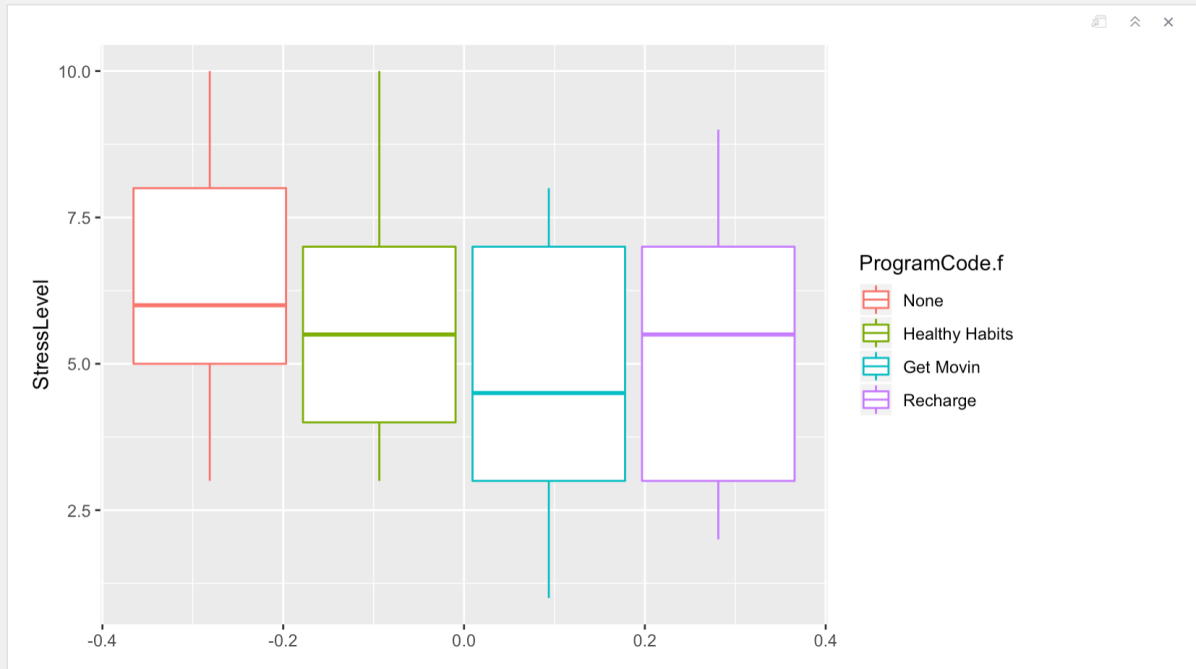   stress level for the Get Movin treatment group was lowest (4.533).

36

```r
37  # Create boxplots of StressLevel across treatment groups
38  ```{r}
39  ggplot(stress, aes(y = StressLevel, group = ProgramCode.f, color = ProgramCode.f)) +
40    geom_boxplot()
41  ```
```


This plot shows how the dependent variable, stres level, varied across the four program groups. The Get Movin group had the lowest maximum, minimum, and median stress levels. Interestingly, the upper quartile for the Healthy Habits group was similar to those for the other two treatments, but this group had a larger maximum stress level and larger lower quartile than the Get Movin and Reharge groups. The boxplot indicates that there weren't outliers in these data.

```r
44  # Conduct an ANOVA
45  ```{r}
46  lm1 <- lm(StressLevel ~ ProgramCode.f, data = stress)
47  anova(lm1)
48  ```
```

```
Analysis of Variance Table

Response: StressLevel
               Df Sum Sq Mean Sq F value  Pr(>F)
ProgramCode.f   3  54.83 18.2750  3.5623 0.01643 *
Residuals     116 595.10  5.1302
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
49  ## Obtain eta squared
50  ```{r}
51  54.83 / (54.83 + 595.10)
52  ```
```

```
[1] 0.08436293
```

```
53
54  The p-value for the model F-test is significant at p<0.05, indicating that there was a significant effect of program on stress
    level.
55  The df-error for this model is 116 (df-err = N-k).
56  The df-hyp for this model is 3 (df-hyp = k-1).
57  The eta squared for the model was 0.084, indicating that the program variable exlained 8.4% of the total variability in stress
    level. Therefore, 91.6% of the total variability in stress level was due to things other than the program variable.
58
```

```r
59  # Conduct planned contrasts
60  ## Create contrasts
61  ```{r}
62  c1 <- c(-1, 1, 0, 0)
63  c2 <- c(-1, 0, 1, 0)
64  c3 <- c(-1, 0, 0, 1)
65  ```
66
67  ## Bind contrasts to predictor variable
68  ```{r}
69  contrasts(stress$ProgramCode.f) <-  cbind(c1,c2, c3)
70  #View the contrasts (At the bottom of the output)
71  stress$ProgramCode.f
72  ```
```

```
  [1] None           Healthy Habits Get Movin       Recharge       Get Movin      Recharge       None
  [8] Recharge       Recharge       Healthy Habits Healthy Habits Healthy Habits Recharge       Healthy Habits
 [15] Recharge       Recharge       Healthy Habits None           Get Movin      Get Movin      Get Movin
 [22] Healthy Habits Healthy Habits Get Movin       Recharge       Get Movin      None           None
 [29] Recharge       Recharge       Get Movin       Get Movin      Healthy Habits Recharge       Get Movin
 [36] Recharge       Healthy Habits None           Healthy Habits Healthy Habits None           Healthy Habits
 [43] Get Movin      Get Movin      Get Movin       Get Movin      Recharge       Healthy Habits Recharge
 [50] Healthy Habits Healthy Habits None           Get Movin      Get Movin      None           None
 [57] Healthy Habits Recharge       Healthy Habits Get Movin      Healthy Habits Recharge       Get Movin
 [64] None           Get Movin      None           None           Healthy Habits Recharge       None
 [71] Recharge       None           None           Get Movin      Healthy Habits None           Get Movin
 [78] Healthy Habits Get Movin      None           Healthy Habits None           Get Movin      Healthy Habits
 [85] Get Movin      None           None           Recharge       None           Get Movin      Recharge
 [92] Healthy Habits Get Movin      Recharge       None           Recharge       Healthy Habits Recharge
 [99] Healthy Habits Recharge       Get Movin      Recharge       None           Healthy Habits Get Movin
[106] Healthy Habits None           None           Recharge       Healthy Habits None           Recharge
[113] Get Movin      Recharge       None           Get Movin      None           Recharge       Recharge
[120] None
attr(,"contrasts")
               c1 c2 c3
None           -1 -1 -1
Healthy Habits  1  0  0
Get Movin       0  1  0
Recharge        0  0  1
Levels: None Healthy Habits Get Movin Recharge
```

73

```r
74 ▾ # Run ANOVA with contrasts
75 ▾ ```{r}
76   lm_contrast <- lm(StressLevel ~ ProgramCode.f, data = stress)
77   summary.lm(lm_contrast)
78   ```
```

```
Call:
lm(formula = StressLevel ~ ProgramCode.f, data = stress)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5333 -1.8333 -0.2667  1.8000  4.1667

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.4750     0.2068  26.479  < 2e-16 ***
ProgramCode.fc1  0.3583     0.3581   1.001  0.31911
ProgramCode.fc2 -0.9417     0.3581  -2.629  0.00971 **
ProgramCode.fc3 -0.2750     0.3581  -0.768  0.44412
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.265 on 116 degrees of freedom
Multiple R-squared:  0.08436,   Adjusted R-squared:  0.06068
F-statistic: 3.562 on 3 and 116 DF,  p-value: 0.01643
```

79
80 The model intercept, 5.475, represents the grand mean of the dependent variable (stress level).
81
82 The planned contrasts revealed that the Get Movin group had significantly lower stress levels compared to the None group ($t(116) = -2.629$, $p < 0.005$ (one-tailed)). The planned constrasts revealed that stress level did not significantly differ between the Healthy Habits and None groups ($t(116) = 1.001$, $p > 0.05$), or between the Recharge and None groups ($t(116) = -0.768$, $p > 0.05$).
83
84 Remember to divide the p value by two when interpreting these effects! The model is two-tailed by default, but planned contrasts are interpreted as one-tailed.
85
86 Planned contrasts are done when you have specific hypotheses that you want to test about relations between certain groups of a predictor variable and the outcome variable. You would use a post-hoc analysis if you didn't have specific hypotheses about differences between groups.